RELIGHTING CHARACTER MOTION FOR PHOTOREAL SIMULATIONS.

Bruce Lamond*, Charles-Felix Chabert, Per Einarsson, Andrew Jones, Wan-Chun Ma, Tim Hawkins , Mark Bolas[‡], Sebastian Sylwan, Paul Debevec
University of Southern California Institute of Creative Technologies, LA, CA, 90292
[‡]University of Southern California Cinema-Television Interactive Media Division, LA, CA 90089

ABSTRACT.

We present a fully image-based approach for capturing and modeling real human locomotion under varying illumination and viewpoint that overviews the techniques and results presented by [Einarsson et al, 2006]. An actor performs repeatable locomotive actions (walking/running) on a rotating treadmill while being filmed from a vertical array of 3 high-speed cameras under controlled rapidly changing lighting conditions. The known rotation of the treadmill, repeatability of the actor's motion, timing of the lighting pattern and capture rate of the cameras are all carefully synchronized so that the actor is imaged in (approximately) the same position in the locomotion at the same point in the lighting pattern but having rotated a known amount due to the known turntable motion. This allows us to effectively multiply the number of cameras from 3 × 1 in azimuth to 3 × 36. Small perturbations in the actor's repeating cyclic position are corrected for using optical flow, and optical flow is also used to align images temporally. This leads to a *flowed reflectance field* data structure. Datasets are compressed using image compression. Image-based relighting and a combination of view morphing and light field rendering implemented on the GPU allow us to render the subject under novel viewpoint and illumination. To composite the person into a scene we derive an alpha matte from retro-reflective material and a back-lit diffuse backdrop, and implement a voxel-based visual hull process to compute how the person should cast shadows on the ground plane. We demonstrate realistic composites of real subjects into real and virtual environments applicable to the area of training simulation.

1. INTRODUCTION.

In the realm of virtually realistic training simulation research, cross-disciplinary factors such as AI, character modeling and animation, and immersive hardware have received the majority of attention so far. Current immersive systems however tend to have more of the look of a contemporary computer game than that of the real world. Character models tend to be blocky with perceptible animation artifacts inconsistent with realistic human articulation. Textures are typically acquired from statically-lit images or hand drawn and, being static, do not display the subtle lighting dynamics that one would expect to see in a real world video sequence of the scene. Our approach to creating realistic character simulations

Figure 1. *Multiple instances of a captured subject rendered into an image-based lighting environment.*

can be classified as post-production control of viewpoint and illumination on predetermined actor performances (Fig. 1). The work presented here overviews our method and results in [Einarsson et al, 2006] within the context of simulation. This data-driven approach circumvents the problems outlined above by working only from real video sequences of subjects. Using only real images of subject means we do not have to worry about creating realistically modeled and animated virtual characters as these characteristics are already contained in the data. Furthermore, because the sequences are captured under a sufficiently dense set of component lighting conditions, these lighting conditions subtly encode the light interactions representative of any novel illumination environment under all but the most specialized of lighting conditions.

Much work in computer graphics has already addressed post-production control of illumination and viewpoint, although most have either addressed only one or the other of these. Post-production control of viewpoint has been achieved from 2 principal directions: using either sparse camera arrays or dense camera arrays. Using a sparse array involves filming the subject from a few cameras and then projecting these images onto approximate geometric models of the subject derived from or fit to the images [Rander et al, 1997; Moezzi et al, 1996; Matusik et al, 2000; Carranza et al, 2003; Vedula et al, 2005]. While these techniques allow a large measure of viewpoint control, they tend to suffer from rather obvious texture misregistration artifacts and limited sampling of the subject's directional reflectance. Using dense arrays involves a large number of cameras and light field rendering [Levoy and Hanrahan, 1996] to interpolate for novel views of the subject. While these techniques do

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Relighting Character Motion For Photoreal Simulations	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Southern California Institute of Creative Technologies, LA, CA, 90292**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**8**	

not require explicit character geometry and produce realistic results, the domain of viewpoint control is determined by the spatial extent of the camera array. None of these viewpoint control techniques attempt to model the scene under alternative lighting.

Control over illumination after-the-fact has been addressed in [Wenger et al, 2005]. They use a sphere of LED light sources and a high speed camera and capture a performance illuminated under roughly 100 lighting directions repeating every 24^{th} of a second. Image-based relighting is used to render the performance into various new lighting environments. Though highly realistic, their work only deals with performances from a single viewpoint and is only for close-up head shots and only for distant lighting. Our work builds on theirs by tackling these limitations. Control over both viewpoint and illumination has been addressed recently in [Theobalt et al, 2005] using a few cameras with fitted geometry and reflectometry. Their results suffer due to the difficulty of producing representative surface reflectance from a single lighting condition, and from imperfect geometry.

1.1 Contributions.

This work takes one step further towards the goal of capturing a subset of real world performances while having complete control over the performance in post-production in terms of illumination and viewpoint. We achieve this by building on the performance relighting technique of [Wenger et al, 2005] but for full-body capture, including local ground plane lighting interactions, and with a novel viewpoint control method derived from a data structure called a *flowed reflectance field*. In order to obtain a moderately dense array of cameras, we restrict our consideration to cyclic motions such as walking or running. By filming repeatable motions on a slowly rotating turntable from a fixed vertical array of 3 cameras, we synchronize the motion, lights, cameras and turntable to effectively view the same pose repetition in the motion from multiple incremental positions of rotation (ie view approximately the same pose from many more positions than we have cameras). In effect we obtain a number of views intermediate to the sparse and dense camera array techniques described previously. We compute optical flow between neighboring viewpoints and use a combination of light field rendering [Levoy and Hanrahan, 1996; Gortler et al, 1996] and view-interpolation [Chen and Williams, 1993; Seitz and Dyer, 1996; Zitnick et al, 2004] to generate novel viewpoint images of the subject. This latter approach improves upon the method of [Wilburn et al, 2005] who use view-interpolation to smoothly move the viewpoint but only within the plane of the camera array. We allow the viewpoint an extra degree of freedom in that it can be moved in 3 dimensions. For compositing, we produce matte images of the subject using a combination of retro-reflective materials and back-lit diffuse backdrop, and we use a voxel-based visual hull to calculate how the subject should cast shadows on the ground plane. We show different subjects composited realistically into both synthetic and real 3D environments.

2. RELATED WORK.

The technique presented here builds on a wealth of previous work in image-based modeling and rendering, in the following areas in particular:

2.1 View Interpolation and Light field Rendering.

These methods allow novel viewpoints to be generated from previously acquired images. [Chen and Williams, 1993] warps rendered images using depth maps to generate novel views. [Laveau and Faugeras, 1994] use stereo correspondence to compute depth in altering the viewpoint in real scenes. [Seitz and Dyer, 1996] presents a view morphing method for creating correct perspective for novel viewpoints between corresponded original views. [Levoy and Hanrahan, 1996; Gortler et al, 1996] synthesize new views of a scene by sampling rays from a dense 2D array of viewpoints, the latter showing that fidelity can be increased by projecting image samples on to scene geometry. [Miller et al, 1998] also explore this increased fidelity with a *surface light field*.

2.2 Dynamic Light field Acquisition.

These methods construct a 2D array of cameras to capture light fields of dynamic events. [Yang et al, 2002] uses distributed rendering to allow multiple viewers to observe virtual views in real-time. [Yu et al, 2002] extends this approach to *surface cameras* where the light field can focus on non-planar geometry. [Zhang and Chen, 2004] uses depth information to focus a real-time light field from a self-reconfigurable camera array. [Wilburn et al, 2005] uses video from a large array of cameras to perform view interpolation between views in both space and time using optical flow.

Our work integrates and extends these techniques by acquiring a moderately sampled 2D array of images surrounding the subject and combines view-interpolation with light field rendering from optical flow to generate views from novel 3D positions. We make use of optical flow maps instead of computing explicit geometry, which allows our method to effectively handle positional discrepancies in similar poses of the performance which is better than if we tried to compute a mesh from this imperfect data. [Buehler et al, 2001] acquires a time-varying light field across multiple cycles of repeating subject motions with a 2D camera array. We build on this approach by constructing a time-varying reflectance field

and construct a 2D camera array from repeating motions of a rotating subject.

2.3 Image-Based Relighting.

These methods simulate novel illumination from a linear combination of images with different basis lighting conditions. Such techniques have been in the context of rendered images [Dorsey et al, 1995; Nimeroff et al, 1994] and human faces [Debevec et al, 2000]. This relightable data can be mapped onto traditional CG models for real-time rendering [Sloan et al, 2002; Ramamoorthi and Hanrahan, 2002]. [Debevec et al, 2000] describes a *non-local reflectance field* as a 6D function $R = R(\Theta_i, \Phi_i; u_r, v_r, \Theta_r, \Phi_r)$ as the space of radiant light fields R_r $(u_r, v_r, \Theta_r, \Phi_r)$ that result from illuminating a subject from the set of distant lighting directions (Θ_i, Φ_i). Sampled datasets of these functions have been used to render novel illumination and views on virtual objects such as trees [Meyer et al, 2001], real objects [Matusik et al, 2002; Matusik et al, 2002a], and faces [Hawkins et al, 2004]. Such techniques do not deal with capturing dynamic scenes or simulate the photometric interaction of the object with its environment. [Wilburn et al, 2005] captures relightable datasets of human performance but does not deal with changing viewpoint.

2.4 Free-Viewpoint Video.

These techniques generate novel views of live-action sequences from a sparse array of cameras mapped onto a basic geometric model of the subject. [Rander et al, 1997] achieves this using stereo correspondence; [Moezzi et al, 1996] uses silhouette intersection; [Matusik et al, 2000] calculates an image-based visual hull; and [Carranza et al, 2003] fits a surface model to image silhouettes. This class of method suffers from texture misalignment due to errors in the recovered geometry, although this can be improved somewhat with view-dependent texture mapping [Debevec et al, 1996]. [Zitnick et al, 2004] exhibits high-quality nocel views of dynamic scenes using a layered representation for stereo correspondence but is limited to motion within a 1D array of cameras. [Vedula et al, 2002; Vedula et al, 2005] recover *scene flow* for a performance by computing geometry from a sparse set of cameras and calculating the movement of 3D surface points against the geometry. This allows renderings of novel views in time and space. Methods in this class also suffer from failing to take account of changing illumination (except [Theobalt et al, 2005] described earlier). Our work builds on these methods by using a flowed light field view interpolation approach which avoids requiring scene geometry. Time-multiplexed lighting also provides rich information to the relighting process. Unfortunately our method is limited to short sequences of repetitive motion due to the requirement for more viewpoints and higher frame rates.

3. APPARATUS.

We designed and built an acquisition stage (Fig. 2) to capture a large number of 2D images of a performance over time (1D), illumination (2D) and viewpoint (2D) giving a 7D dataset. The focal point has a subject performing on a treadmill placed atop a rotating turntable. The treadmill belt and turntable top are covered with Reflecmedia Chromatte retro-reflective cloth used in the matting process. Shallow channels have been cut into the board under the treadmill belt to give the subject a tactile reference for remaining centered. We use a general-purpose lighting apparatus modified from [Wenger et al, 2005]. The device is the top 2/3 of an 8m diameter 6th-frequency geodesic sphere designed to hold an optimal distribution of 901 controllable light sources. Where [Wenger et al, 2005] captures a working volume of around 50cm diameter, this apparatus has been designed to have a working volume of 2m diameter for human-sized capture. Each light source consists of 6 LumiLEDs Luxeon V LEDs arranged in an 18cm diameter hexagon. Each LED uses a Fraen 'single wide' optic delivering 100 lux to the center of the stage 4m away. Lighting bases used in this work comprise an average of 40 lights, allowing well-exposed images to be captured at 990fps and f2.8. The lights are controlled by 75 microcontroller boards running at 40MHz based on Microchip's PIC 18F8627. A master controller sends a global sync pulse to drive the lighting sequence and trigger the high-speed cameras; it also controls an audible metronome to indicate walk cycle pace to the subject.

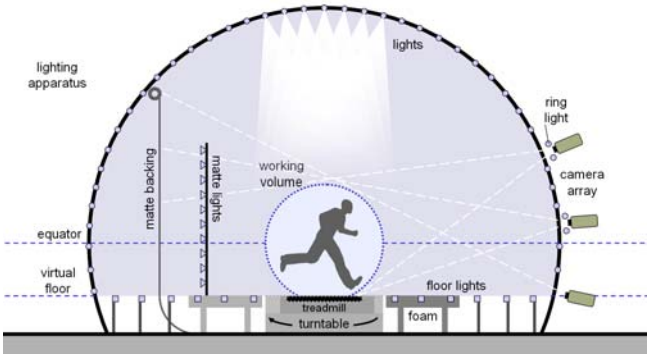


Figure 2. *A side view schematic of our capture stage.*

Another departure from [Wenger et al, 2005] is the stage's 140 floor lights. These consist of 6 optic-less LEDs each in a linear pattern placed at the height of the turntable, 85cm below the stage equator, to simulate illumination from a Lambertian ground plane beneath the subject. Small vertical mirror pairs are oriented behind each LED to increase the amount of light cast towards the subject with increasing distance of the light from the subject. Dome and floor light intensities are calibrated by acquiring lighting basis images of a 30cm 33% gray sphere in a few positions in the working volume.

We image the subject with a vertical array of 3 Vision Research Phantom 7.1 high-speed digital video cameras placed just outside the dome, one at the virtual floor level and the other two at 17º and 34º above the floor. The top two cameras each have a ring of 6 LEDs placed just around the lens and fitted with Fraen 'single narrow' optics aimed along the camera axis. These are used to illuminate the retro-reflective cloth and matting backdrop for compositing the subject. The matting backdrop is a 3m x 4m sheet of 18% gray paper behind the subject, illuminated by 2 stands of 22 additional light sources in a vertical array behind and to the sides of the camera fields.

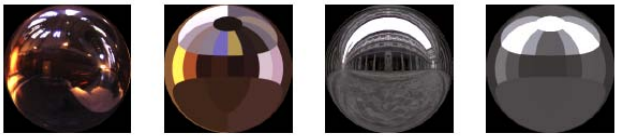
4. ACQUISITION.

To capture a performance the natural walk/run rate for the subject is recorded. The treadmill and turntable speeds are then set accordingly to give us 36 cycles of motion in 360º of rotation (Fig 3b). We activate the lighting and once the subject is moving comfortably we begin filming. We capture the performance with repeating sets of illumination conditions at 30 sets per second (chosen to fit to 30fps video frame rates). Each set consists of 33 lighting conditions: 26 lighting direction bases, 3 each of evenly spaced tracking frames and matting frames, and 1 unused stripe pattern (Fig 3a). The 32 used lighting frames are similar in function to [Wenger et al, 2005]. The small number of conditions reflects a tradeoff we have had to make: we want to capture 36 cycles of 320 × 448 pixels within each cameras limited memory capacity of 8GB. Improvements in camera technology will likely alter this tradeoff favorably. The 26-element lighting basis is chosen to represent a real-world environment in a small number of conditions and to be symmetrical with the vertical axis as we have to rotate the lighting environment with the person's angle when rendering. The resolution is finer in the upper hemisphere (23 conditions) with just 3 for the ground plane (Fig. 4). The 36 locomotion cycles recorded by 3 cameras yields 108 relightable cycles. After capture, a *clean plate* sequence of the setup without the actor is obtained; geometric calibration is found from imaging a human-sized checkerboard using [Zhang, 2000]; photometric calibration is done by imaging a MacBeth ColorChecker chart.

5. GENERATING THE FLOWED REFLECTANCE FIELD.

After capturing a subject, we obtain an alpha channel for the images and compute optical flow to register the images spatially and temporally. Finally the images are compressed into the flowed reflectance field.

Figure 3. *(a) 1 set of basis lighting conditions taken from the middle camera in 1/30th sec. The sequence shows 26 basis lighting conditions, 3 matte and track frames (and 1 unused stripe pattern). (b)The 36×3 array of viewpoints for a single pose in the cycle.*

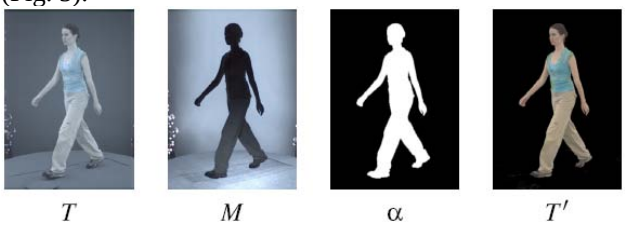


(a) (b) (c) (d)

Figure 4. *(a,c) 2 lighting environments and (b,d) their projections onto the 26-element basis.*

5.1 Computing Mattes.

An alpha channel [Porter and Duff, 1984] or matte is generated for each tracking frame T_0, T_1, and T_2 in each lighting set. The alpha channel is derived from the track frame T and the matte frame M following each track frame. During a matte lighting condition, the main lights for the other lighting directions are turned off and the special matting lights are turned on to light the retro-reflective cloth and matting backdrop. To compute the matte we compare neighboring track and matte frames and retain inferred foreground pixels from the track frame whose monochrome brightnesses are greater then in the matte frame. We then eliminate stray foreground elements from the track frame by excluding pixels not part of the largest central connected matte component and apply a 1-pixel Gaussian blur to model the filtering introduced by the image sensor and color interpolation processes to yield the final matte α. The pre-multiplied foreground T' is computed by matting T onto a black background using the clean plate image C and α using $T' = T - (1 - \alpha).C$ (Fig. 5).



T M α T'

Figure 5. *Matting. A tracking frame T, consecutive matte frame M, alpha matte image α, and T matted onto black and color-corrected T'*

5.2 Lighting Basis Registration.

We now register the 33 images in a lighting set temporally following [Wenger et al, 2005]. We wish to warp the 33 images so that they are aligned with the tracking frame in the middle of the sequence. This yields sharper images in the image-based relighting. Optical flow vectors are calculated using the algorithm from [Black and Anandan, 1993] from the middle tracking frame T_1 to the other tracking frames T_1 and T_2 after matting the track frames to reduce image clutter and maximize the robustness of the flow process. We then interpolate flow between the tracking frames to warp the other lighting frames to the T_1 using a reverse pixel lookup. Lighting frames outwith T_0 and T_2 are warped by extrapolating the flow slightly since no frame is more than $1/60^{th}$ second from the central frame and little artifacts are observed in the rectification. From here the matte for T_1 can now be used as the matte for the co-aligned sequence of 33 frames and the sequence is matted onto black using α and *C*.

5.3 Flow Between Viewpoints.

Each pose in the locomotion has a 36 × 3 grid of 4D reflectance fields $R_{u,v}(s)$ where (u,v) is the horizontal and vertical viewpoint index and s is the 2D image coordinate in that view. To create the flowed reflectance field, optical flow between each viewpoint and its 4 neighbors is computed (images on the top and bottom rows have only 3 flow fields). These flow fields are denoted $F^{\rightarrow}_{u,v}(s)$ where the arrow indicates the direction of the image toward which the flow has been computed. Flow fields are stored relative to s so that s in reflectance field $R_{u,v}$ corresponds to pixel coordinate $(s + F^{\rightarrow}(u,v,s))$ in reflectance field $R_{u+,v}$. If there is no motion between 2 images, the flow field is zero.

Bidirectional flow between neighboring pairs of vertical viewpoints is first computed. Since the views are acquired from different cameras, the image $R_{u,v+}$ is projected onto the frontoparallel plane through the origin as viewed by reference camera $R_{u,v}$ to produce a warped field $R'_{u,v+}$. We then compute the corresponding pixel coordinate in the warped image for each pixel in the reference image by projecting the coordinate through the inverse of the warping homography. Bidirectional flow between neighboring horizontal image pairs are captured from consecutive walk cycles and thus do not show the subject in exactly the same position. We widen the search space to compute this horizontal flow, although we can omit the homography rectification process since image pairs are captured in the same camera. Fig. 6 shows a visualization of a flowed reflectance field.

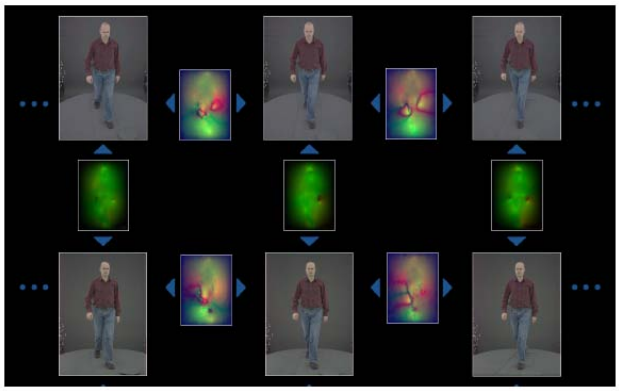


Figure 6. *Bi-directional optical flow maps for neighboring viewpoints in the dataset. Up/down displacement is green and left/right is red.*

5.4 Computing Shadows.

Our shadow computation technique models a first order approximation of how the subject should interact with the ground plane to add a qualitative sense of realism to the composite. To compute the shadows we model a 3m × 3m plane below the subject consisting of 128^2 pixels. Volumetric intersection of space [Szeliski, 1993] defined from a number of mattes of a pose give us an approximate visual hull of the subject. Rays are traced from each ground pixel towards the light sources to determine the percentage of light from each basis condition that remains visible. The result is a set of 23 attenuation maps (one for each lighting basis in the upper hemisphere) for each pose in the locomotion (Fig. 7).

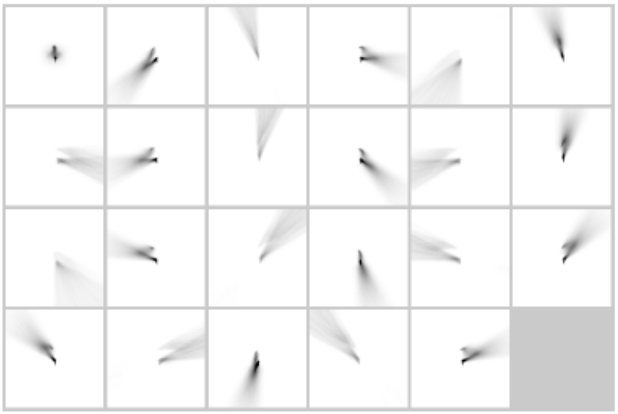


Figure 7. *Shadow maps for one frame in a walk cycle for the 23 basis lighting conditions above ground.*

5.5 Compression.

A single data capture accumulates 3 × 8GB of 12-bit raw data in our cameras. During the 6 hours of processing, we compress the data to approximately 1.5GB of reflectance fields and 1GB of flow maps. The flow maps use quantized pixel displacements using a signed logarithmic scale that is able to hold pixel flow variations

in 8-bits. Quantized maps are then further compressed using Huffman encoding to 34% of original size. To compress the reflectance fields, we compress the images rather than the reflectance functions, since the lighting resolution is relatively coarse. Images are JPEG compressed using mosaics of the lighting basis images in gamma 2.2 corrected space. Images are combined according to relighting coefficients with floating point accuracy after decompressing and applying gamma. This process achieves 16:1 compression with only very minor artifacts.

6. RENDERING.

The rendering process consists of re-lighting, image warping, light field interpolation, shadow rendering and compositing.

6.1 Relighting.

Each reflectance field is first lit using an image-based environment. The environment must first be oriented to match the rotation of the subject for that field. The environment is projected onto the lighting bases to produce image-based relighting coefficients. The flowed reflectance field is now a flowed light field consisting of a set of 36 × 3 arrays of pre-lit images, one for each pose in the motion. Each pre-lit image forms a vertex of a squashed cylindrical polygon.

6.2 Warping and Light Field Interpolation.

The rendering process computes morphs between images in the dataset according to interpolation coefficients calculated from a light field rendering process. We present: linear interpolation by a scalar β between a particular pre-lit image and its right-hand neighbor is:

$$I'_{u,v,\beta}(s) = \bar{\beta} I_{u,v}(s) + \beta I_{u^+,v}(s)$$

Where $\bar{\beta} = 1 - \beta$ and $u^+ = u + 1$. Morphing between the 2 images based on their flow maps is similarly:

$$I'_{u,v,\beta}(s) = \bar{\beta} I_{u,v}(s + \beta F^{\leftarrow}_{u^+,v}(s)) + \beta I_{u^+,v}(s + \bar{\beta} F^{\rightarrow}_{u,v}(s))$$

The process is shown in Fig. 8. The displacement of the pixel coordinate sampled from image $I_{u,v}$ is taken from the flow map $F^{\leftarrow}_{u+,v}$ from the *other* image $I_{u+,v}$ to $I_{u,v}$ and vice-versa. This is so that as β approaches 1, the sample from $I_{u,v}$ approaches the sample that corresponds to $I_{u+,v}$'s pixel at s. The process can be generalized to morph between the 4 vertices of a quad in the squashed cylinder with coefficients β and γ as follows:

$$\begin{aligned} I'_{u,v,\beta,\gamma}(s) = \quad & \bar{\beta}\bar{\gamma} I_{u,v}(s + \beta\bar{\gamma} F^{\leftarrow}_{u^+,v}(s) + \beta\gamma F^{\downarrow}_{u,v^+}(s)) \quad + \\ & \bar{\beta}\bar{\gamma} I_{u^+,v}(s + \bar{\beta}\bar{\gamma} F^{\rightarrow}_{u,v}(s) + \beta\gamma F^{\downarrow}_{u^+,v^+}(s)) \quad + \\ & \bar{\beta}\gamma I_{u,v^+}(s + \beta\gamma F^{\leftarrow}_{u^+,v^+}(s) + \bar{\beta}\bar{\gamma} F^{\uparrow}_{u,v}(s)) \quad + \\ & \beta\gamma I_{u^+,v^+}(s + \bar{\beta}\gamma F^{\rightarrow}_{u,v^+}(s) + \beta\bar{\gamma} F^{\uparrow}_{u^+,v}(s)) \end{aligned}$$

This warping process allows us to generate novel views of the subject from anywhere on the cylindrical viewing surface. If we had a dense sampling of views within all of the cylinder quads, we could use traditional light field rendering to re-bin rays from this surface and generate arbitrary views from 3D positions including points inside and outside of the cylinder. In fact this is exactly how our rendering algorithm works except that we avoid having to generate the dense sampling of views by computing only the pixels s of the morphed views *I'* comprising the final rendered pixels as follows: for each pixel t in novel view *V*:

- Cast a ray *R* through t to intersect the cylinder at point *p* on polygon (*u*,*v*)
- Determine the bilinear interpolation coefficients β, γ corresponding to *p*'s position within the polygon
- Set $V(t) = I'_{u,v,\beta,\gamma}(s)$ where s is the pixel in the image plane of *I'* intersected by ray *R* through *p*

The last step requires that we can infer the intrinsic and extrinsic camera parameters corresponding to a virtual view. We do this by bilinearly interpolating the known parameters from the polygon vertices. Warped matte images are produced using the same procedure.

We have implemented flowed light field interpolation on the GPU using OpenGL. The 36 × 3 RGBA images and 180 flow maps for a single pose can be held in 256MB of GPU memory and rendered interactively. To render an animated sequence the images and flow maps for each virtual camera are sent to the GPU. The polygonal geometry of the cylinder is then computed for each pixel. We then pass the texture coordinates and ray intersection point to the fragment shader where virtual camera parameters are inferred along with interpolation coefficients used to warp and blend the contribution from the 4 closest input images.

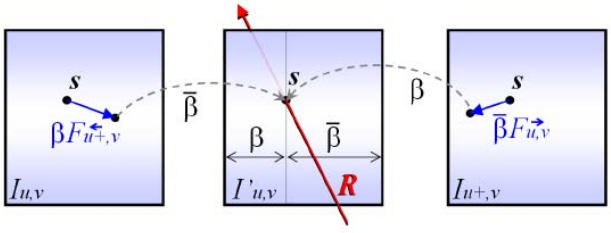


Figure 8. *Novel view generation in a flowed light field*

6.3 Shadowing and Compositing.

Shadows are rendered by applying a similar image-based relighting process to the corresponding shadow map basis for each frame of the animation. This produces

a re-lit shadow map indicating the relative irradiance below the subject. The rendering of the subject is composited over the background scene using α and the *over* operator [Porter and Duff, 1984]

7. RESULTS.

We have captured 3 subjects using our method: a female walking and a male walking and running. Fig. 9(a-c) shows the male walking composited into the captured *Uffizi Gallery* environment. The distant diffuse lighting environment means we can light him with the same environment for the whole sequence. The subject and shadows are composited in a virtual camera move across a high-res image of the environment. Subtle lighting effects can be seen in the subjects skin and the shadows as the camera moves in 3 dimensions.

Fig 9(d-f) shows the female subject rendered into a virtual environment computed using global illumination. In each frame she is lit with varying illumination from omnidirectional HDR images from the position of her torso (inset). As she moves through the scene, various dominant lighting effects can be seen along with subtle indirect reflections from the colored walls.

Fig 9(g-i) shows the male running composited in another image-based environment. The environment was captured near sunset giving rich warm indirect lighting from the building reflections.

Fig 1 shows multiple instances of the male running in another image-based environment. Although individual instances cast shadows on the ground, the instances do not interact with each other. This is left as future work.

CONCLUSIONS.

We have presented a new method for creating image-based renderings for a subset of human motions with control in post-production over lighting and viewpoint. The cyclic nature of our motions means that we have been able to acquire a 6D reflectance field from a 1D array of cameras. We have been able to interpolate and extrapolate moderately spaced positions in our data using optical flow. The technique shows notably improved realism compared to previous image-based approaches. Further iterations of this class of technique will undoubtedly produce extremely realistic renditions of human performances for use in training simulations.

ACKNOWLEDGEMENTS

This work was sponsored by the University of Southern California Office of the Provost and the U. S. Army Research, Development, and Engineering Command (RDECOM). The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred

REFERENCES.

BLACK, M. J., AND ANANDAN, P. 1993. A framework for the robust estimation of optical flow. In *Fourth International Conf. on Computer Vision*, 231–236.

BUEHLER, C., BOSSE, M., McMILLAN, L., GORTLER, S. J., COHEN, M. F. 2001. Unstructured Lumigraph Rendering. In *Proceedings of ACM SIGGRAPH 2001* (Aug 2001), Computer Graphics Proceedings, Annual Conference Series, pp. 425-432.

CARRANZA, J., THEOBALT, C., MAGNOR, M. A., AND SEIDEL, H.-P. 2003. Free-viewpoint video of human actors. *ACM Transactions on Graphics 22*, 3 (July), 569–577.

CHEN, S. E., AND WILLIAMS, L. 1993. View interpolation for image synthesis. In *Proceedings of SIGGRAPH 93*, Computer Graphics Proceedings, Annual Conference Series, 279–288.

DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. 1996. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of SIGGRAPH 96*, Computer Graphics Proceedings, Annual Conference Series, 11–20.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000* (July), 145–156.

DORSEY, J., ARVO, J., GREENBERG, D. 1995. Interactive Design of complex time dependent lighting. *IEEE Computer Graphics and Applications* 15, 2 (Mar. 1995), 26-36

EINARSSON, P., CHABERT, C-F., JONES, A., MA, W-C., LAMOND, B., HAWKINS, T., BOLAS, M., SYLWAN, S., DEBEVEC, P. 2006. Relighting Human Locomotion with Flowed Reflectance Fields. In *Eurographics Symposium on Rendering: 17th Eurographics Workshop on Rendering.*

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *Proceedings of SIGGRAPH 96*, Computer Graphics Proceedings, Annual Conference Series, 43–54.

HAWKINS, T., WENGER, A., TCHOU, C., AND DEBEVEC, A. G. F. G. P. 2004. Animatable facial reflectance fields. In *Eurographics Symposium on Rendering: 15th Eurographics Workshop on Rendering*.

LAVEAU, S., AND FAUGERAS, O. 1994. 3-D scene representation as a collection of images. In *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, 689–691.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *Proc. of SIGGRAPH 96*, Computer Graphics Proceedings, Annual Conference Series, 31–42.

MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S. J., AND MCMILLAN, L. 2000. Image-based visual hulls. In *Proc. SIGGRAPH 2000*, 369–374.

MATUSIK, W., PFISTER, H., NGAN, A., BEARDSLEY, P., ZIEGLER, R., AND MCMILLAN, L. 2002. Image-based 3d photography using opacity hulls. *ACM Transactions on Graphics* 21, 3 (July), 427–437.

MATUSIK, W., PFISTER, H., ZIEGLER, R., NGAN, A., AND MCMILLAN, L. 2002a. Acquisition and rendering of transparent and refractive objects. In *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering*, 267–278.

MCMILLAN, L., AND BISHOP, G. 1995. Plenoptic modeling: An image-based rendering system. In *Proceedings of SIGGRAPH 95*, Computer Graphics Proceedings, Annual Conference Series, 39–46.

MEYER, A., NEYRET, F., AND POULIN, P. 2001. Interactive rendering of trees with shading and shadows. In *Rendering Techniques 2001: 12th Eurographics Workshop on Rendering*, 183–196.

MILLER, G. S. P., RUBIN, S., AND PONCELEON, D. 1998. Lazy decompression of surface light fields for precomputed global illumination. *Eurographics Rendering Workshop 1998* (June), 281–292.

MOEZZI, S., KATKERE, A., KURAMURA, D. Y., AND JAIN, R. 1996. Reality modeling and visualization from multiple video

sequences. *IEEE Computer Graphics & Applications* 16, 6 (Nov.), 58–63.

NIMEROFF, J. S., SIMONCELLI, E., AND DORSEY, J. 1994. Efficient re-rendering of naturally illuminated environments. In *5th Eurographics Workshop on Rendering*, 359–373.

PORTER, T., AND DUFF, T. 1984. Compositing digital images. In *Computer Graphics (Proceedings of SIGGRAPH 84)*, vol. 18, 253–259.

RAMAMOORTHI, R., AND HANRAHAN, P. 2002. Frequency space environment map rendering. *ACM Transactions on Graphics 21*, 3 (July), 517–526.

RANDER, P., NARAYANAN, P. J., AND KANADE, T. 1997. Virtualized reality: Constructing time-varying virtual worlds from real events. In *Proceedings of IEEE Visualization*, 277–283.

SEITZ, S. M., AND DYER, C. R. 1996. View morphing: Synthesizing 3d metamorphoses using image transforms. In *Proceedings of SIGGRAPH 96*, Computer Graphics Proceedings, Annual Conference Series, 21–30.

SLOAN, P.-P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics 21*, 3 (July), 527–536.

SZELISKI, R. 1993. Rapid octree construction from image sequences. *CVGIP: Image Understanding 58*, 1 (July), 23–32.

THEOBALT, C., AHMED, N., DE AGUIAR, E., ZIEGLER, G., LENSCH, H., MAGNOR, M., AND SEIDEL, H.-P. 2005. Joint motion and reflectance capture for creating relightable 3d videos. Technical Report MPI-I-2005-4-004, Max-Planck-Institut fuer Informatik.

VEDULA, S., BAKER, S., KANADE, T. 2002. Spatio-temporal view interpolation. In *Rendering Techniques 2003: 13th Eurographics Workshop on Rendering* (June 2002), pp. 65-76.

VEDULA, S., BAKER, S., KANADE, T. 2005. Image-based spatio-temporal modeling and view interpolation of dynamic events. *ACM Transactions on Graphics* 24, 2 (Apr. 2005), 240-261.

WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics* 24, 3 (Aug.), 756–764.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High performance imaging using large camera arrays. *ACM Transactions on Graphics* 24, 3 (Aug.), 765–776.

YANG, J. C., EVERETT, M., BUEHLER, C., AND MCMILLAN, L. 2002. A real-time distributed light field camera. In *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering*, 77–86.

YU, J., McMILLAN, L., GORTLER, S. 2002. Scan light field rendering. In *Pacific Graphics* (Beijing, China, Oct. 2002).

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics* 23, 3 (Aug.), 548–558.

ZHANG, Z. 2000. A flexible new technique for camera calibration. *PAMI* 22, 11, 1330–1334.

ZITNICK, C. L., KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2004. High-quality video view interpolation using a layered representation. *ACM Transactions on Graphics* 23, 3 (Aug.), 600–608.

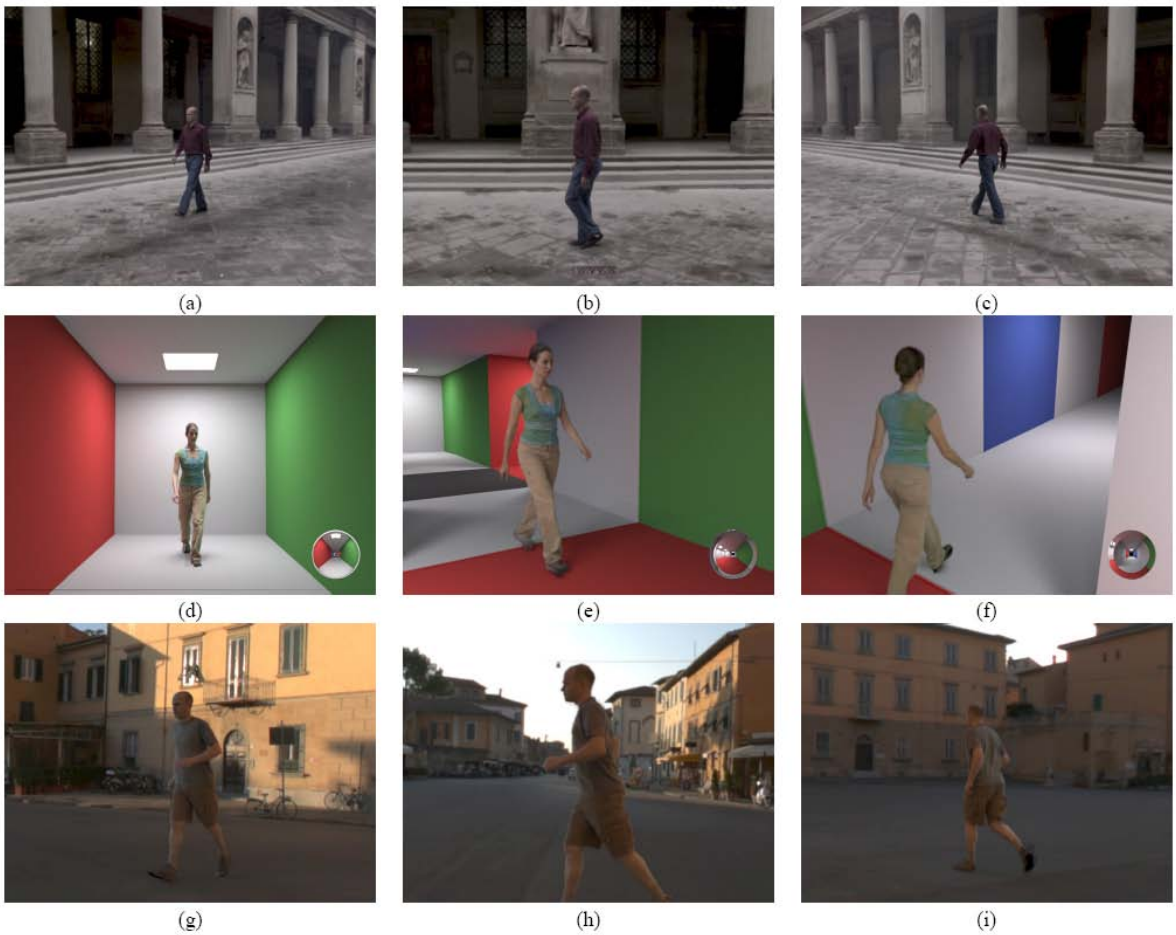


Figure 9. *Results. (a-c) walking male subject composited into image-based 'Uffizi' environment. (d-f) female subject added to virtual scene with global illumination effects. (g-i) running male composited into a different image-based lighting environment.*